IMM-NYU 334
FEBRUARY 1965

NEW YORK UNIVERSITY
COURANT INSTITUTE OF
MATHEMATICAL SCIENCES

OPTIMAL ROBUSTNESS FOR ESTIMATORS AND TESTS

ALLAN BIRNBAUM AND EUGENE LASKA

PREPARED UNDER
CONTRACT NONR 285(38)
NR 042-206/11/18/64

IMM-NYU-334
February 1965

New York University

Courant Institute of Mathematical Sciences

OPTIMAL ROBUSTNESS FOR ESTIMATORS AND TESTS

Allan Birnbaum

Eugene Laska*

This report represents results obtained at the Courant Institute of Mathematical Sciences under the sponsorship of the Office of Naval Research under Contract No. Nonr 285(38). Reproduction in whole or in part is permitted for any purpose of the United States Government.

*Research Facility, Rockland State Hospital, Orangeburg, N. Y.

1. Introduction and summary. Tukey (1960,1962) has provided a broad perspective for research in efficiency-robustness of estimators, as well as an important part of the knowledge available in this area. The present paper is intended to complement these by supplying formulations of concepts, techniques, and initial results for optimally efficiency-robust estimators and tests in several types of problems. Relations to Tukey's investigation are discussed in Section 2, with brief reference to the related work of Huber (1964). Relations to the approach to robust estimation of Hodges and Lehmann (1963) are discussed in Section 3.

The present approach may be described as a formal indexing of alternative specifications (e. g. "shapes" of error-distributions) by a nuisance parameter, and adaptation of admissibility and related concepts and Bayes techniques of the Neyman-Pearson and Wald theories to the estimation and testing problems thus formulated. Specific problems for which new optimal efficiency-robust estimators are given are: linear estimation of location parameters (Section 2); rank tests and related estimators for two-sample problems (Section 3); and unbiased estimation (Section 4). A by-product included in Section 4 is a generalization of Stein's (1950) characterization of locally-best unbiased estimators to the class of admissible unbiased estimators together with the corresponding complete class theorem.

2. Linear unbiased estimation of location parameters. Let X be a random variable with p.d.f. $f((x-\mu)/\sigma,\lambda)$, where the finite variance σ^2, mean μ, and shape parameter λ are unknown but have the specified ranges $-\infty < \mu < \infty$, $\sigma^2 > 0$, $\lambda \in \Lambda$. For each λ we assume that the density function is symmetric. Let $(X_1,\ldots,X_n)$ denote n independent observations on X, and let $Y = (Y_1,\ldots,Y_n)$ denote the same observations ordered nondecreasingly. We consider the problem of estimation of μ, restricting consideration to linear unbiased estimators (LUEs), that is, estimators of the form $\mu^* = \sum_i^n a_i y_i$ for which $E(\mu^*(Y)|\mu,\sigma,\lambda) \equiv \mu$ identically in μ, σ, and $\lambda \in \Lambda$. Estimators will be appraised in terms of their variance functions $\mathrm{var}(\mu^*|\mu,\sigma,\lambda)$. When Λ consists of a single point (i. e. the shape is known), the problem is reduced to one solved by Lloyd (1952), who derived best linear unbiased estimators (BLUEs) (of σ as well as μ, without the assumption of symmetry made here). With λ unknown, the problem leads to considerations which are conveniently illustrated first in the artificially simple case that Λ contains just two elements, and σ is known. For example the two shapes might be normal and double-exponential,

$$f((x-\mu),1) = \frac{1}{\sqrt{2\pi}} \exp - \tfrac{1}{2}(x-\mu)^2, \quad f((x-\mu),2) = \tfrac{1}{2} \exp - |x-\mu|,$$

or normal and 5%-contaminated normal as defined in Tukey's work.

2.1 Illustrative discussion. Among the estimators of possible interest in any such simplified problem, let us consider initially the BLUE estimators, to be denoted by $\hat{\mu}_\lambda$, $\lambda = 1$ or 2, and their

variances $\mathrm{var}(\hat{\mu}_\lambda | \lambda')$, $\lambda, \lambda' = 1,2$, which turn out to be independent of μ and σ (as seen in Lloyd's results of their extension below). Since these are BLUE estimators under respective shapes, we have $\mathrm{var}(\hat{\mu}_1|1) \leq \mathrm{var}(\hat{\mu}_2|1)$ and $\mathrm{var}(\hat{\mu}_2|2) \leq \mathrm{var}(\hat{\mu}_1|2)$. If the first relation were an equality, $\hat{\mu}_2$ would be a uniformly best linear unbiased estimator (UBLUE) over Λ, and would provide a simple and ideal solution to our problem. In cases of typical interest all such inequalities are strict, and more complicated considerations must be faced. Several such cases are illustrated in Figures 1-4, in which each estimator μ^* is represented by its variance function $\mathrm{var}(\mu^*|\lambda)$, $\lambda = 1,2$, plotted as the point with the latter as respective coordinates. (These considerations are in part analogous to familiar decision-theoretic discussions of the convex set of "α, β" points in the problem of testing between two simple hypotheses.) In the hypothetical example of Figure 1, BLUE $\hat{\mu}_1$ has variance function $(1,4)$, and BLUE $\hat{\mu}_2$ has variance function $(6,2)$. Thus use of $\hat{\mu}_1$ risks a four-fold increase of variance in case $\lambda = 2$ is the true shape, and use of $\hat{\mu}_2$ risks a three-fold increase of variance in case $\lambda = 1$ is true. Since each of these estimators is uniquely best under one shape, each alternative estimator is represented by a point of the form $(1+a, 2+b)$ with both a and b positive; the general goal is to seek estimators for which both a and b are small.

<u>Definitions</u> (not restricted to case of two-point Λ): Among LUEs, μ^* is naturally called <u>better than</u> μ^{**} if $\mathrm{var}(\mu^*|\lambda) \leq \mathrm{var}(\mu^{**}|\lambda)$

over Λ with strict inequality for at least one λ. μ^* is called an admissible linear unbiased estimator (ALUE, with respect to Λ) if no other estimator is better.

Under certain general but not universal conditions attention may be restricted without loss to admissible estimators, in the sense that for each inadmissible estimator there exists at least one better admissible estimator. It is proved below that the risk points of the ALUEs constitute a convex curve connecting the risk points of the BLUEs, as in Figures 1-3.

Depending upon the structure of a problem such as that represented in Fig. 1, the ALUEs might be represented by convex curve passing very near the "ideal" point (1,2); or by one lying very near the line-segment connecting the BLUE points as in Fig. 2. In the first case any estimator with risk point very near (1,2) would naturally be called highly efficiency-robust since it has efficiency nearly 100% uniformly over Λ. In the second case, it is seen that uniformly high efficiency is unattainable, but that efficiency at least 54% is provided by an ALUE with risk point near (1.8, 3.68); here BLUE $\hat{\mu}_1$ provides efficiency bounded below by only 50%, and $\hat{\mu}_2$ by only 1/6. In problems having a third type of structure, one BLUE estimator may have efficiency nearly 100% uniformly over Λ; in such a case further consideration of other ALUEs would be of minor practical importance. In such comparisons of estimators, it may be of interest to consider the formal criterion of maximin efficiency (the maximum attainable lower bound of efficiency of a LUE over Λ); or the criterion of minimax

Figure 1

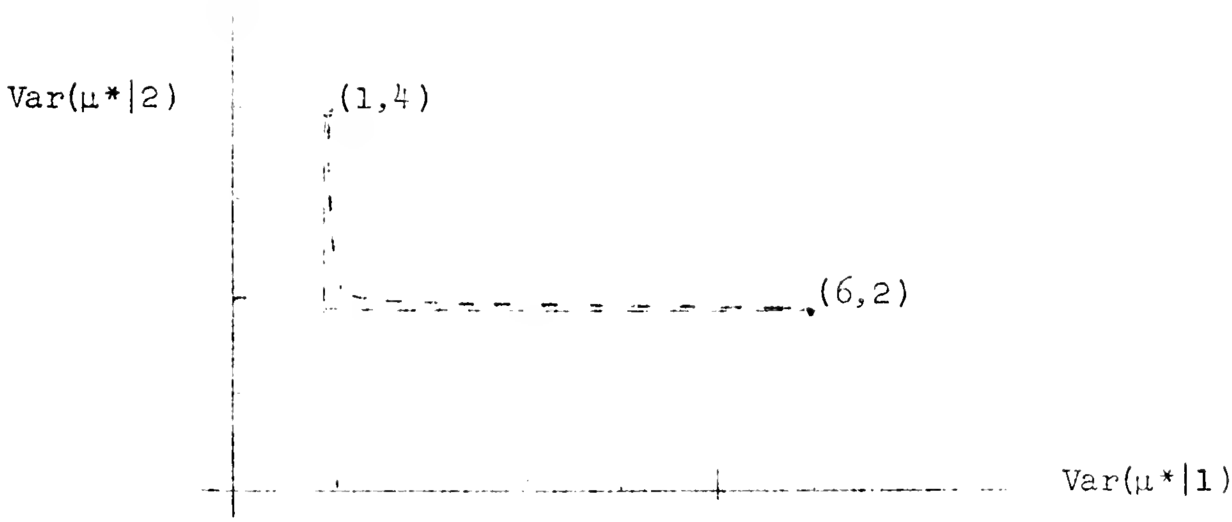
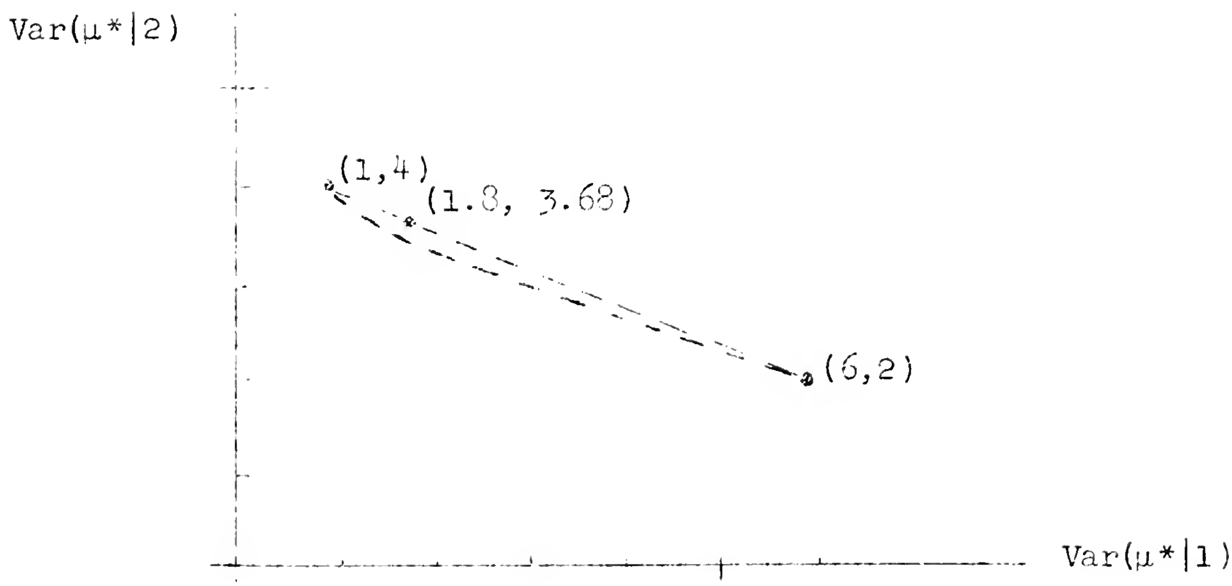


Figure 2



variance (in problems formulable with some common scale unit for different shapes). However the three types of structures of simplified problems described serve to illustrate the limited value of any single criterion or measure of efficiency-robustness, and the fact that practical interest leads in principle to consideration of (a) the configuration of risk points of BLUE estimators; and in many cases also (b) the configuration of risk points of ALUE estimators.

While problems in which Λ contains only several points are artificially simplified versions of more realistic problems, it may be good research strategy to begin efficiency-robustness investigation of a realistic model Λ by preliminary consideration of one or several such simplified versions. If for example two shapes are selected for preliminary study, from among those in a model of practical interest; and if a structure like that of Figure 2 is found, in which maximin efficiency is far below unity; then it follows that in the original problem embracing additional shapes the maximin efficiency can be no higher (in general it will be lower). Such negative results found early in an investigation may provide economical redirection of research efforts. On the other hand, positive results found in simplified problems provide tentative encouragement and suggest specific estimators worth considering further in the original problem. (Hypothetical examples can be seen by retrospective consideration of the results in Tukey (1960), Figure 4, p. 460, considering pairs of shapes such as $\gamma = 0$ or $.05$. For each such pair, the estimators considered can be represented in a figure analogous to Figures 1-3.)

The body of knowledge of efficiency-robustness available at any stage can be viewed usefully in terms of the headings:

(a) Negative results: pairs, or larger sets, Λ of shapes over which uniformly high efficiency is unattainable.

(b) Positive results: sets of shapes Λ over which one can attain certain fairly high lower bounds on efficiency, together with construction of corresponding estimators. (The 5% truncated mean represented in Fig. 4 of Tukey (1960) is a significant example.)

(c) Intermediate cases Λ.

Of course practical judgments concerning the range Λ, and concerning gains or sacrifices in terms of variances or efficiencies at various λ when estimators are compared, are crucial to practically useful formulations of problems and selections of estimators. Tukey's papers introduce a specific "contaminated normal" family of distributions as a model pertinent to many practical problems of estimation of location and scale, together with discussion of the background of experience and theory which suggest such a model. (Cf. also Buhler (1964).)

The class of estimators (of location and scale) investigated by Tukey for possible robustness of <u>asymptotic</u> efficiency under this model were LUEs. It had been shown that restriction to such estimators does not generally reduce asymptotic efficiency in cases of known shapes, (Blom (1958)); and considerations described by Tukey focused attention particularly on certain of these (truncated (or trimmed) or Winsorized means for location). Evidently the attainability of uniformly high efficiency of estimation of location, even in an unrestricted class of estimators, was not

guaranteed *a priori* by any theoretical knowledge available, but it was found that certain truncated means were uniformly highly asymptotically efficient over the given Λ. In this case there would be little to be gained by considering ALUEs over the given class; more precisely, for specific finite sample sizes for which Tukey's asymptotic results effectively hold, there would be little to gain by considering other estimators - but in other cases here, and more generally in other investigations, the construction of some ALUEs and computations of their variances can be considered. (Huber's (1964), pp. 74-5, comments on the relative merits of variances and asymptotic variances are pertinent here.)

2.2 Derivations. Let $U_r = (X_r-\mu)/\sigma$ and $V_r = (Y_r-\mu)/\sigma$, for $r = 1,\ldots n$. The moments of the standardized ordered observations V_r are independent of μ and σ but in general depend upon λ: Let

$$\alpha_r^\lambda = E(V_r|\mu,\sigma,\lambda), \quad \text{and} \quad \omega_{r,s}^\lambda = \operatorname{Cov}(V_r,V_s|\mu,\sigma,\lambda)$$

for $r,s = 1,\ldots n$. Then for the original ordered observations we have

$$E(Y_r|\mu,\sigma,\lambda) = \mu + \sigma\alpha_r^\lambda, \quad \text{and} \quad \operatorname{Cov}(Y_r,Y_s|\mu,\sigma,\lambda) = \sigma^2\omega_{r,s}^\lambda.$$

Lloyd's method of constructing BLUE estimators can be applied here to construct ALUE after application of the usual formal Bayes technique: Let $G = G(\lambda)$ denote an arbitrary cumulative

probability distribution function over Λ, and let

$$f((x_1-\mu)/\sigma, \ldots (x_n-\mu)/\sigma, G) = \int_\Lambda \prod_{i=1}^{n} f((x_i-\mu)/\sigma, \lambda) dG(\lambda).$$

The latter is the p.d.f. of a sample $(X_1, \ldots X_n)$ taken under one or another shape λ, with λ assigned values randomly according to distribution G. We have

$$E(Y_r|\mu,\sigma,G) = \int_\Lambda E(Y_r|\mu,\sigma,\lambda) dG(\lambda)$$

$$= \int_\Lambda (\mu+\sigma\alpha_r^\lambda)\ dG(\lambda) = \mu + \sigma\alpha_r^G ,$$

where
$$\alpha_r^G = \int_\Lambda \alpha_r^\lambda\ dG(\lambda),$$

and

$$Cov(Y_rY_s|\mu,\sigma,G) = E(Y_rY_s|\mu,\sigma,G) - (\mu+\sigma\alpha_r^G)(\mu+\sigma\alpha_s^G)$$

$$= \int [\sigma^2\omega_{rs}^\lambda + (\mu+\sigma\alpha_r^\lambda)(\mu+\sigma\alpha_s^\lambda)]dG(\lambda) - (\mu+\sigma\alpha_r^G)(\mu+\sigma\alpha_s^G)]$$

$$= \sigma^2 \int \omega_{rs}^\lambda dG(\lambda) + \mu^2 + \sigma^2 \int \alpha_r^\lambda\alpha_s^\lambda dG(\lambda)$$

$$+ \mu\sigma \int (\alpha_r^\lambda+\alpha_s^\lambda)dG(\lambda) - \mu^2$$

$$- \sigma^2 \int \alpha_r^\lambda dG(\lambda) \int \alpha_s^\lambda dG(\lambda) - \mu\sigma \int (\alpha_r^\lambda+\alpha_s^\lambda)dG(\lambda)$$

$$= \sigma^2 \left[\int \omega_{rs}^\lambda dG(\lambda) + \int \alpha_r^\lambda\alpha_s^\lambda dG(\lambda) - \int \alpha_r^\lambda dG(\lambda) \int \alpha_s^\lambda dG(\lambda) \right]$$

$$= \sigma^2\omega_{r,s}^G, \text{ say.}$$

Thus the conditions of the Gauss-Markoff theorem are satisfied, for minimum-variance linear unbiased estimation of μ, when the values of α_r^G and $\omega_{r,s}^G$ are known.

Let $y = (y_1,\ldots y_n)'$, $\omega^G = (\omega_{r,s}^G)$, $D_G = (\omega^G)^{-1}$, $\omega^\lambda = (\omega_{r,s}^\lambda)$, $\alpha^G = (\alpha_1^G,\ldots\alpha_n^G)'$, $\alpha^\lambda = (\alpha_1^\lambda,\ldots\alpha_n^\lambda)'$, $\underline{1} = (1,1,\ldots 1)'$, an $(n\times 1)$ vector, $p = (\underline{1},\alpha^G)$, and $\theta = (\mu,\sigma)'$. Then

$$E(Y|\mu,\sigma,G) = p\theta \quad\text{and}\quad \mathrm{Cov}(Y|\mu,\sigma,G) = \sigma^2\omega^G.$$

The best (minimum variance) unbiased estimator linear in the y_r's is, by the Gauss-Markoff theorem,

$$\hat{\theta}_G = (p'D_Gp)^{-1}p'D_Gy$$

and the covariance matrix of the estimator is

$$\mathrm{Cov}(\hat{\theta}_G|\mu,\sigma,G) = \sigma^2(p'D_Gp)^{-1}$$

$$= \sigma^2\begin{bmatrix} \underline{1}'D_G\,\underline{1} & \underline{1}'D_G\alpha^G \\ \underline{1}'D_G\alpha^G & \alpha^{G'}D_G\alpha^G \end{bmatrix}^{-1}.$$

In the important case of $f((x-\mu)/\sigma,\lambda)$ symmetric about 0 for each λ, to which our further discussion and results are limited, we have more simply

$$\hat{\theta}_G = [\underline{1}'D_Gy/\underline{1}'D_G\underline{1},\ \alpha^{G'}D_Gy/\alpha^{G'}D_G\alpha^G]$$

and

$$\mathrm{Cov}(\hat{\theta}_G|\mu,\sigma,G) = \sigma^2\begin{bmatrix} 1/\underline{1}'D_G\,\underline{1} & 0 \\ 0 & 1/\alpha^{G'}D_G\alpha^G \end{bmatrix} ;$$

and D_G and ω^G are symmetric; and $\alpha_i^\lambda = -\alpha_{n-i}^\lambda$ for $i = 1,\ldots n$. Restricting consideration now to estimation of μ we find

$$E(\mu_G|\lambda) = \underline{1}'D_G(\mu\underline{1}+\alpha^\lambda)/\underline{1}'D_G\underline{1} = \mu.$$

Thus for each G, $\hat{\mu}_G$ is an unbiased linear estimator of μ under each shape $\lambda \in \Lambda$.

For our purposes each such estimator is represented by its variance function $(\mathrm{Var}(\hat{\mu}_G|\mu,\sigma,\lambda) =$

$$\sigma^2[\underline{1}'D_G\omega^\lambda D_G'\,\underline{1}/(\underline{1}'D_G\,\underline{1})^2],$$

which is independent of μ. This function of λ may be interpreted as a "risk point" characterizing $\hat{\mu}_G$, in a space with coordinate axes indexed by the respective $\lambda \in \Lambda$ and with corresponding coordinates given by the preceding variance function, as in the simplified examples of Figures 1-3 above. To prove that each $\hat{\mu}_G$ is an admissible linear unbiased estimator, it suffices to note that it is a unique Bayes solution for the G-mixture. Hence μ_G uniquely minimizes $\mathrm{Var}(\mu^*|\mu,\sigma,G)$ over a class of linear estimators which includes all those unbiased for each λ. If $\hat{\mu}_G$ were inadmissible, there would exist a LUE μ' with $\mathrm{Var}(\mu'|\mu,\sigma,\lambda) \leqq \mathrm{Var}(\hat{\mu}_G|\mu,\sigma,\lambda)$ for each λ and hence with

$\mathrm{Var}(\mu'|\mu,\sigma,G) \leq \mathrm{Var}(\hat{\mu}_G|\mu,\sigma,G)$, contradicting that $\hat{\mu}_G$ uniquely attains the latter minimum.

To prove that the set of risk points $\mathrm{Var}(\mu_G|\mu,\sigma,\lambda)$, $\lambda \in \Lambda$, of such admissible LUE's constitutes a convex hypersurface (or convex curve as in the simplified examples of Figures 1-3) consider for each pair of ALUE's μ^*, μ^{**} and each number q, $0 \leq q \leq 1$, the "mixed" estimator μ': for each y, $\mu'(y)$ is assigned the value $\mu^*(y)$ with probability q and the value $\mu^{**}(y)$ with probability $1-q$. Clearly μ' is unbiased and has risk point

$$\mathrm{Var}(\mu'|\mu,\sigma,\lambda) = q\ \mathrm{Var}(\mu^*|\mu,\sigma,\lambda) + (1-q)\ \mathrm{Var}(\mu^{**}|\mu,\sigma,\lambda).$$

The latter formula shows that the risk points of all mixed LUE's constitute a convex set, which includes the set of risk points of ALUE's. Next, let us consider again the "Bayes risk", to be minimized as above but with respect to this larger class of "mixed" estimators: We have

$$\begin{aligned}\mathrm{Var}(\mu'|\mu,\sigma,G) &= \int \mathrm{Var}(\mu'|\mu,\sigma,\lambda)dG(\lambda)\\ &= q\int \mathrm{Var}(\mu^*|\mu,\sigma,\lambda)dG(\lambda) + (1-q)\int \mathrm{Var}(\mu^{**}|\mu,\sigma,\lambda)dG(\lambda)\\ &= q\ \mathrm{Var}(\mu^*|\mu,\sigma,G) + (1-q)\ \mathrm{Var}(\mu^{**}|\mu,\sigma,G).\end{aligned}$$

Thus the problem of minimizing the latter function in the class of "mixed" estimators μ' is seen to reduce to the problem solved above, and to be answered by the (non-mixed) ALUE $\hat{\mu}_G$. Thus the

latter has a risk point which is a "lower" boundary point of the convex set referred to above. Arguments of continuous variation in the class of distributions G (which are elementary in the case illustrated in Figures 1-3) show that the set of such risk points not only lies in, but constitutes, the convex lower boundary hypersurface of that set.

Example. If Λ contains just two points, $\lambda = 0,1$, any G may be represented by $g = \text{Prob}\ [\lambda=0]$, $1-g = \text{Prob}\ [\lambda=1]$, and

$$f((x-\mu)/\sigma, g) = gf((x-\mu)/\sigma, 0) + (1-g)\ f((x-\mu)/\sigma, 1).$$

$$E(y) = \mu+\sigma(g\alpha^{0}+(1-g)\alpha^{1})$$

$$\text{Cov}(YY') = \sigma^{2}[g\omega' + (1-g)\omega^{0} + g(1-g)(\alpha^{1}-\alpha^{0})(\alpha^{1}-\alpha^{0})^{1}].$$

It is readily found that for $n = 3$, $\hat{\mu}_g = \sum_i^3 a_i y_i$ with

$$a_1 = a_3 = \frac{g(\omega_{22}^{0}-\omega_{12}^{0}) + (1-g)(\omega_{22}^{1}-\omega_{12}^{1})}{g(\frac{1}{2}\omega_{11}^{0} + \frac{1}{2}\omega_{13}^{0} + \omega_{22}^{0} - 2\omega_{12}^{0}) + (1-g)} \cdot (\tfrac{1}{2}\omega_{11}^{1} + \tfrac{1}{2}\omega_{13}^{1} + \omega_{22}^{1} + 2\omega_{12}^{1})$$

and $a_2 = 1-2a_1$.

Here, as might be expected in this very simple case, the $\hat{\mu}_g$'s are respective weighted averages of $\hat{\mu}_0$ and $\hat{\mu}_1$. It is not clear whether this relation holds for larger n and for more general Λ.

If the two shapes are normal and double-exponential,

$$f((x-\mu),0) = \frac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}(x-\mu)^2\right],$$

$$f((x-\mu),1) = \tfrac{1}{2} \exp\left\{-|x-\mu|\right\},$$

we have the ALUE's and their risk points given in Table 1 and Figure 3. The moments from which these were calculated are available in Hastings, et al (1947) and Sarhan (1954).

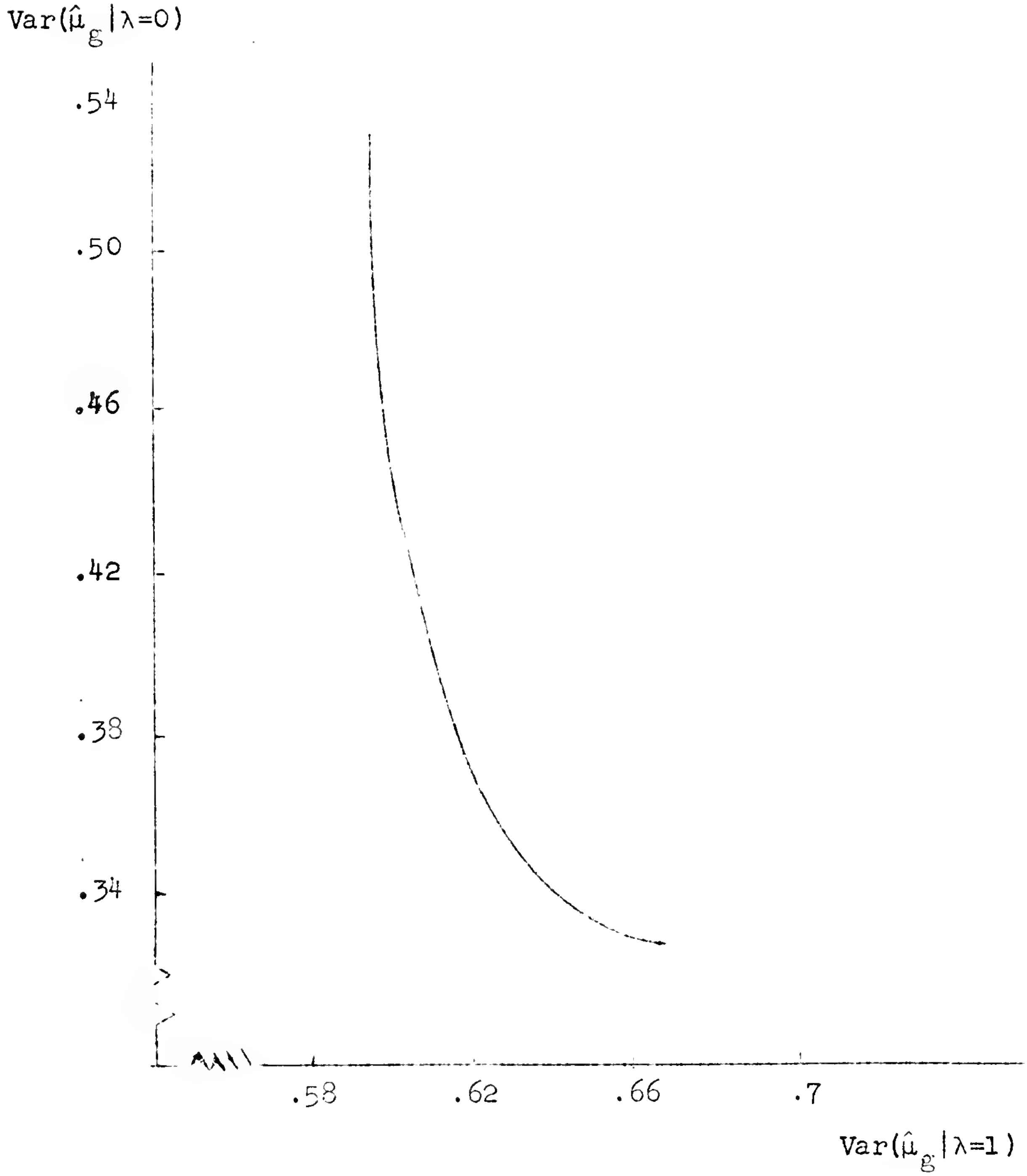
TABLE 1

g	$a_1 = a_3$	a_2	$var(\mu_g \mid \lambda=0)$	$var(\mu_g \mid \lambda=1)$
0	.148	.704	.532	.591
.25	.226	.548	.407	.605
.50	.275	.450	.351	.628
.75	.309	.382	.343	.649
1.00	.333	.333	.333	.668

Here, as could be seen from the risk points of just $\hat{\mu}_o$ and $\hat{\mu}_1$, use of $\hat{\mu}_o$ (or else another ALUE with a nearby risk point) is recommended by the fact that achievement of any major part of the small decrease possible in $Var(\hat{\mu}_g \mid \lambda=1)$ costs a large increase in $Var(\hat{\mu}_g \mid \lambda=0)$.

Figure 3

RISK POINTS OF A COMPLETE CLASS OF
ADMISSIBLE LINEAR UNBIASED ESTIMATES



For any given family of symmetric shapes $f((x-\mu)/\sigma,\lambda)$, $\lambda \in \Lambda$, the properties of ALUE's μ_g are characterized by their respective variance functions $^2[\underline{1}'D_G\omega^\lambda D_G' \underline{1} / (\underline{1}'D_G \underline{1})^2]$. Use of these functions, and the formulae for the estimators themselves,

$$\hat{\mu}_G = \underline{1}'D_G y/\text{tr } D_G ,$$

depend upon availability of the moments $\alpha^\lambda, \omega^\lambda$ for respective shapes and sample sizes of interest; and upon feasibility of the indicated matrix computations for at least several distributions G in each problem to be explored. Availability of such moments for shapes representing plausible error-distributions is increasing (e. g. Greenberg and Sarhan (1962), Birnbaum and Dudman (1963)); but for such important cases as the contaminated normal family moments are not available. The computation of $D_G = (\omega^G)^{-1}$ from given α^λ's, ω^λ's, and G is evidently generally heavy, even when Λ contains only several points; conceivably these might be facilitated by discovery of algebraic simplifications and use of large scale computers.

It will often be natural to consider families of shapes which are convex, in the sense that for each q, $0 < q < 1$, if $\lambda', \lambda'' \in \Lambda$ then the shape

$$qf((x-\mu)/\sigma,\lambda') + (1-q)f((x-\mu)/\sigma,\lambda'')$$

is also in Λ. The contaminated normal family, for example, has this property, which was also assumed in the investigation of

Bühler (1964). (The property clearly cannot hold when Λ contains a finite number of points.) An interesting open question is whether, for such convex families Λ, each ALUE μ_G is also a BLUE $\hat{\mu}_\lambda$ for some $\lambda \in \Lambda$.

3. Efficiency-robust two-sample rank tests and estimators.

3.1 Introduction. Let $x_1, x_2, \ldots, x_{n_1}, y_1, y_2, \ldots, y_{n_2}$ be $N = n_1 + n_2$ independent random variables, the x_i's having common unknown absolutely continuous c.d.f. P_X and the y_i's having common unknown absolutely continuous c.d.f. P_Y. Let $\frac{\partial G}{\partial \theta}(x, \theta) = g(x, \theta)$ be a specified one-parameter family of density functions with θ taking values in an open interval. We consider the problem of testing the simple hypothesis

$$H_o: \quad P_X = G(x, \phi_o), \; P_Y = G(y, \phi_o)$$

against a composite alternative

$$H_1: \quad P_X = G(x, \theta), \quad P_Y = G(y, \phi)$$

where θ and ϕ are any values in R for which $\theta > \phi_o > \phi$.

It is well known that the tests which are valid (i. e. which have specified size under H_o for all continuous G) are just the rank tests, those based on the rank order statistics

$$z_{Nj} = \begin{cases} 1 & \text{if the } j^{th} \text{ smallest of the pooled sample of } N \text{ observations is an } x, \\ 0 & \text{otherwise.} \end{cases}$$

Such nonparametric two-sample tests might be described as "universally validity-robust". In sharpest contrast with the breadth of this validity property, the efficiency theory of such tests has been restricted exclusively to considerations of power under one or another assumed known form for G. In the present section we extend this theory systematically to the case of $G(x, \theta, \lambda)$ where θ remains the parameter of primary interest and where the (nuisance) "shape" parameter λ has an unknown value in a specified set Λ. To illustrate by reference to familiar rank tests, it is well known that the Fisher test is uniquely locally-best if G is normal, and that the Wilcoxon test is uniquely locally-best if G is logistic. Thus use of either test risks a certain loss of efficiency in case that shape holds for which it is not best. Hence it is of much theoretical and practical interest to ask what other tests should be considered here, with a view to attaining the highest possible efficiencies under the two possible shapes. Introductory and interpretive remarks here will be few because they would be for the most part direct analogues of Section 2.1 above, but some comments will be made here with particular reference to rank tests, and in Section 3.3 below with reference to corresponding estimators.

It has sometimes been found that LMPRT has some efficiency-robustness properties (e. g. the Wilcoxon test, LMPRT for logistic shapes, is also highly efficient for normal shapes). While such results are valuable when found, it seems of practical value for other efficiency-robustness problems, and of general theoretical value, to provide systematic theory and techniques for efficiency-

robustness in this area.

One type of question, which is rather basic to the comparison of parametric and non-parametric approaches in general, is illustrated by the following particular results (due to Chernoff and Savage (1958) and Capon (1961). The Fisher rank test is known to be asymptotically efficient, as compared with the two-sample t-test, under normal shapes; and analogous efficiency holds for the Wilcoxon test, as compared with the parametric likelihood-ratio test under logistic shapes, as it does for all LMPRTs compared to the corresponding likelihood ratio test. Further, the Fisher test is asymptotically preferable to the t-test in the sense that its efficiency relative to the latter is at least unity for <u>all</u> shapes (under mild assumptions); however other LMPRTs fail to have the analogous property: There are shapes under which the relative efficiency of the LMPRT to its parametric analogue falls below unity. In such appraisals of given rank tests, for relative efficiency in relation to respectively parametrically-best tests for various families of shapes, it would seem natural and promising to extend consideration to rank tests characterized by optimal efficiency-robustness.

3.2 <u>Efficiency-robust rank tests</u>. Capon (1961), making use of a general theorem of Hoefding, has shown that, under certain regularity conditions, the locally most powerful rank test (LMPRT) is given by a critical function of the form

$$\phi(z) = \begin{cases} 0 & \text{if} \quad T(z_{N1},\ldots,z_{NN}) < C_\alpha \\ 1 & \text{if} \quad T(z_{N1},\ldots,z_{NN}) > C_\alpha \\ k_\alpha & \text{if} \quad T(z_{N1},\ldots,z_{NN}) = C_\alpha \end{cases}$$

where

$$T_N = T_N(z) = T(z_{N1},\ldots,z_{NN}) = \frac{1}{n_1}\sum_{i=1}^{N} a_{Ni}z_{Ni}$$

$$a_{Ni} = E_{\phi_o\phi_o}\left[\frac{\partial}{\partial\theta}\log g(W_i,\theta)\bigg|_{\theta=\phi_o}\right]$$

and where W_i is the i^{th} smallest of the N observations. The notation $E_{\theta\phi}(\cdot)$ indicates expectation is taken under the assumption that the x's and y's are distributed according to $g(x,\theta)$ and $g(y,\phi)$ respectively.

The size α_{T_N} of the test is given by

$$\alpha_{T_N} = \sum_{z} \phi(z)P\{z|H_o\}$$

and the power function $\beta_{T_N}(\theta,\phi)$ by

$$\beta_{T_N}(\theta,\phi) = \sum_{z} \phi(z)P\{z|H_1\}$$

where $P\{z|H_i\}$ represents the probability that $Z = z$ under H_i, $i = 0,1,$ and the summation is over all possible z.

We turn now to the case in which the form of G is incompletely known; and to the corresponding problem of constructing admissible rank tests (ARTs)(with respect to specified families Λ of shapes), with attention again focused on a neighborhood of ϕ_o. Let $\Omega = \{(\theta,\lambda)\}$ represent a specified class of c.d.f.'s $G(x,\theta,\lambda) = G_\theta(x,\lambda)$, where the range of θ is an open interval and $\lambda \in \Lambda$, a specified family of "shapes". For any test δ, we shall refer to the negative of the derivative of the power function,

$$-\beta_\delta'(\phi_o,\phi_o,\lambda) = -\frac{\partial}{\partial\phi}\beta_\delta(\phi_o,\phi,\lambda)\Big|_{\phi=\phi_o}$$

as the <u>risk function</u> $r = r(\delta,\lambda)$, $\lambda \in \Lambda$.

Let $\mathcal{L}_\alpha$ be the class of rank tests $\delta(z)$ such that the probability of a type I error $\beta_\delta(\phi_o,\phi_o,\lambda) = \alpha$ uniformly in λ. Admissibility within $\mathcal{L}_\alpha$ is defined in the usual way in terms of the risk function $r(\delta,\lambda)$. We consider next the problems of characterizing the admissible tests and of examining their risk points.

Let $J(\lambda)$ be any <u>a priori</u> distribution over Λ. We call any test $\delta_J \in \mathcal{L}_\alpha$ a <u>Bayes solution</u> with respect to J if δ_J minimizes the <u>Bayes risk</u>

$$r(\delta,J) = -\int_\Lambda \beta_\delta'(\phi_o,\phi_o,\lambda)dJ(\lambda)$$

over all $\delta \in \mathcal{L}_\alpha$. For any fixed J, consider the family of distributions $G(x,\theta,J) = \int G(x,\theta,\lambda)dJ(\lambda)$ and our testing problem concerning θ applied to this family. We have for any δ

$$r(\delta, J) = \int r(\delta, \lambda) dJ(\lambda) = - \sum_{z} \delta(z) \left[\int_{\Lambda} \frac{\partial}{\partial \phi} \Pr \{z | \phi_o, \phi, \lambda\} \Big|_{\phi=\phi_o} dJ(\lambda) \right]$$

which by use of Hoefding's theorem reduces to

$$r(\delta, J) = \sum_{z} \delta(z) T_N^J(z),$$

where

$$T_N^J(z) = \int_{\Lambda} T_N^{\lambda}(z) dJ(\lambda).$$

Thus $r(\delta, J)$ is minimized essentially uniquely, among all tests in $\mathscr{D}_\alpha$, by

$$\delta_J(z) = \begin{cases} 0 & \text{if} \quad T_N^J > C_\alpha \, , \\ 1 & \text{if} \quad T_N^J < C_\alpha \, , \\ k_\alpha & \text{if} \quad T_N^J = C_\alpha \, . \end{cases}$$

Here "essentially unique" means that any other test minimizing $r(\delta, J)$ has the same risk function as δ_J. The admissibility of each such δ_J follows immediately from the well known property of admissibility of essentially unique Bayes solutions.

The formula for T_N^J above reduces to

$$T_N^J(z) = \frac{1}{n_1} \sum_{i=1}^{N} a_{Ni}^J z_{Ni} \, ,$$

where

$$a_{Ni}^{J} = \int_{\Lambda} a_{Ni}^{\lambda} dJ(\lambda) .$$

This formula admits the interesting and useful interpretation that each of the ARTs δ_J is based on a linear combination of the z_{Ni}'s, resembling the LMPRTs in this respect; and that the "scores" a_{Ni}^{J} defining δ_J are respectively simply the J-weighted averages of the corresponding scores a_{Ni}^{λ}, $\lambda \in \Lambda$. For example, if J gives equal weights $\frac{1}{2}$ to the normal and logistic shapes, then δ_J is given by scores $a_{Ni}^{J} = \frac{1}{2} a_{Ni}^{1} + \frac{1}{2} a_{Ni}^{2}$, where a_{Ni}^{1} is a "normal" (Fisher test) score and a_{Ni}^{2} is a Wilcoxon test score, $i = 1, \ldots N$.

Because each ART δ_J has this form, it is possible to apply familiar techniques to compute its asymptotic efficiency under each shape $\lambda \in \Lambda$. As is the with LMPRTs, for ARTs it is only asymptotic approximations to power functions which are now practically available in numerical form. These are the subject of the following subsection 3.

It is useful in connection with investigations of specific problems to note that the set of risk functions $r(\delta^J, \lambda)$, $\lambda \in \Lambda$, of the respective ARTs constitute a subset of the "lower" boundary hypersurface of the convex set of risk functions of all tests in $\mathcal{L}_\alpha$. The convexity of the latter set follows by a familiar elementary argument from the fact that $\mathcal{L}_\alpha$ is a convex set of critical functions and the observation that $r(\delta, \lambda)$ is a linear functional of δ. Under general but not universal conditions, of which a simple case is that Λ contains a finite number of points, it is known that the admissible class is a complete class in $\mathcal{L}_\alpha$, and

that therefore its risk functions constitute the full "lower" boundary hypersurface mentioned. Evidently a corresponding convexity property holds for asymptotic relative efficiency functions.

3.3 Asymptotic Relative Efficiencies of ARTs. It is convenient, and represents no essential restriction of methods, to present derivations and examples here for the case in which Λ contains two points only, $\lambda = 1,2$. We assume here that

(i) $g_\theta(x,\lambda)$ and $\frac{\partial}{\partial\theta} g_\theta(x,\lambda)$ are continuous with respect to θ in R for almost all x, and $g_\theta(x,\lambda)$ and $\left|\frac{\partial}{\partial\theta} g_\theta(x,\lambda)\right|$ are dominated by functions which are Riemann integrable on the real line,

(ii) $g_\theta(x,\lambda)$ and $g_\phi(x,\lambda)$ have the same support,

(iii) $|J^{(i)}_{G_\lambda}(H)| = \left|\frac{d^i JG_\lambda}{dH^i}\right| \leqq K_\lambda H(1-H)^{-i-\frac{1}{2}+\delta_\lambda}$ for $0 \leqq H \leqq 1$,

$\lambda = \lambda_1, \lambda_2$ (except perhaps for a finite number of H where $J^{(i)}_{G_\lambda}$ may fail to exist) for some $\delta > 0$, k_λ a constant and J_{G_λ} defined by

$$J_{G_\lambda}(G_{\phi_o}(x,\lambda)) = \frac{\partial}{\partial\theta} \ell_n\, g_\theta(x,\lambda)\Big|_{\theta=\phi_o}$$

and where

$$0 < \lim_{N \to \infty} \frac{n_1}{n_2} = r < \infty .$$

Let $G_\theta(x,\lambda_1) = G_\theta(x)$ and $G_\theta(x,\lambda_2) = G^*_\theta(x)$.

Also let

$$J_p = pJ_G + (1-p)J_{G*} \quad 0 \leq p \leq 1 .$$

<u>THEOREM</u>: Under assumptions (i), (ii) and (iii),

$$\lim_{N \to \infty} P_G \left\{ \frac{T_N^p - E_{\Theta\phi}(T_N^p)}{\sigma_{\Theta\phi}(T_N^p)} \leq t \right\} = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

where

$$E_{\Theta\phi}(T_N^p) = \int_{-\infty}^{\infty} J_p(H_{\Theta\phi}(x))dG_\Theta(x)$$

and

$$N\sigma_{\Theta\phi}^2(T_N^p) = \frac{2n_2}{N} \left\{ \iint_{-\infty<x<y<\infty} G_\phi(x)(1-G_\phi(y))J_p'(H_{\Theta\phi}(x))J_p'(H_{\Theta\phi}(y))dG_\Theta(x)dG_\Theta(y) \right.$$

$$\left. + \frac{n_2}{n_1} \iint_{-\infty<x<y<\infty} G_\Theta(x)(1-G_\Theta(y))J_p'(H_{\Theta\phi}(x))J_p'(H_{\Theta\phi}(y))dG_\phi(x)dG_\phi(y) \right\}$$

and where

$$H_{\Theta\phi}(x) = \frac{n_1}{N} G_\Theta(x) + \frac{n_2}{N} G_\phi(x)$$

providing $\sigma_{\Theta\phi}(T_N^p) \neq 0$. In particular under the null hypothesis for G or G^*, $\sigma^2_{\phi_o\phi_o} \equiv \sigma^2_{\phi_o}$ is given by

$$\frac{n_1 N}{n_2} \sigma^2_{\phi_o}(T^p_N) = \int_o^1 J^2_p(x)dx - \left(\int_o^1 J_p(x)dx \right)^2 .$$

This result is an immediate consequence of a theorem of Chernoff and Savage (1958). Ansari and Bradley (1960) have pointed out that if a finite number of exceptional points are allowed in condition (iii), the theorem remains valid.

Obviously the theorem remains true if we replace $G_\theta(x)$ by $G^*_\theta(x)$ above. We use this result to calculate ARE's of T^p_N and ART's.

For purposes of investigating the asymptotic properties of the admissible tests T^p_N we use the Pittman (1948) Noether (1955) criterion for the ARE of two sequences of tests $W = \{W_N\}$ and $W^* = \{W^*_N\}$. Consider the sequence of alternatives $\Delta_N = \theta_N - \phi_N = kN^{-\frac{1}{2}}$ where k is a non-zero constant, $\frac{\phi_N - \phi_o}{\theta_N - \phi_o} = -\frac{n_1}{n_2}$ and $\Delta = \theta - \phi$. Let

$$\lim_{N \to \infty} \frac{n_1}{n_2} = \lim_{N \to \infty} \frac{n^*_1}{n^*_2} = r.$$

Suppose that W and W^* have the same size. The ARE of W to W^* is defined to be $\lim_{N \to \infty} \frac{N^*}{N}$ where N^* is the sample size of the second test required to achieve the same power for a given alternative Δ_N which W_N achieves with respect to the same alternative when using a sample of N observations.

Consider such a sequence of alternatives with $\Delta = \theta - \phi$ and a sequence of statistics $W = \{W_N\}$ and $W^* = \{W^*_N\}$ satisfying the

following three conditions in the neighborhood of $\Delta = 0(\Phi=\Theta=\Phi_o)$:

(a) $$\lim_{N \to \infty} P_F \left\{ \frac{W_N - E_{\Theta\Phi}(W_N)}{\Theta\Phi(W_N)} \leq t \right\} = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

(b) $$\lim_{N \to \infty} \frac{\tilde{\Theta}_N \Phi_N (W_N)}{\tilde{\Phi}_o (W_N)} = 1$$

(c) $$E_W(F) = \lim \left\{ \frac{\frac{\partial}{\partial\Delta} E_{\Theta\Phi}(W_N)}{\left(\frac{n_1 n_2}{N}\right)^{\frac{1}{2}} \tilde{\Phi}_o(W_N)} \right\}^2_{\substack{\Theta = \Phi = \Phi_o \\ \Delta = 0}}$$

exists and is independent of k.

$E_W(F)$ is called the efficacy of W at F and F refers to the distribution under which all expectations above are taken.

Pittman has shown that if (a), (b) and (c) are satisfied for $W = \{W_N\}$ and $W^* = \{W^*_N\}$ then

$$\lim \frac{N^*}{N} = \frac{E_W(F)}{E_{W^*}(F)} \equiv E_{W,W^*}(F)$$

(where $E_{W,W^*}(F)$ is the ARE of W to W*) if $E_{W^*}(F) \neq 0$ and $\lim_{N \to \infty} \frac{n_1}{n_2} = \lim \frac{n^*_1}{n^*_2} = r$.

Consider now $T^p = \{T^p_N\}$. From the theorem we have that condition (a) is satisfied. Capon (1961) has shown that for LMPRT's condition (b) holds and the same argument shows that the same is true for T^p_N.

To check condition (c) we calculate the relevant efficacies. We evaluate

$$\frac{n_1 N}{n_2}\,\sigma^2_{\phi_0}(T_N^p) = \int_0^1 J_p^2(x)dx - \left(\int_0^1 J_p(x)dx\right)^2.$$

The last integral gives

$$\int_0^1 J_p(x)dx = p\int_0^1 J_G(x)dx + (1-p)\int J_{G*}(x)dx.$$

Now

$$\begin{aligned}\int_0^1 J_G(x)dx &= \int_0^1 J_G(G)dG \\ &= E_{\phi_0}\left(\frac{\partial}{\partial\theta}\log\,(x)\Big|_{\phi_0} G\right) \\ &= 0\end{aligned}$$

This last result follows from an interchange of integration and differentiation which may be carried out under the assumptions of the theorem. Similarly we find

$$\int_0^1 J_{G*}(x)dx = 0$$

Thus

$$\frac{n_1}{n_2} N \sigma^2_{\phi_o}(T^p_N) = \int_0^1 J^2_p(x)dx$$

$$= p^2 \int_0^1 J^2_G(x)dx + 2p(1-p) \int_0^1 J_G(x)J_{G*}(x)dx + (1-p)^2 \int_0^1 J^2_{G*}(x)dx.$$

Substituting $x = G_{\phi_o}(y)$ in the first integral and $x = G^*_{\phi_o}(y)$ in the last integral, we get finally

$$\sigma^2_{\phi_o}(T^p_N) = \frac{n_2}{n_1 N} \left\{ p^2 \inf G_{\phi_o} + 2p(1-p) \int_0^1 J_G(x)J_{G*}(x)dx + (1-p)^2 \inf G^*_{\phi_o} \right\}$$

where $\inf G_{\phi_o}$ is R. A. Fisher's "information" of G_θ evaluated at $\theta = \phi_o$. Notice that this is the variance of T^p_N under both G_{ϕ_o} and $G^*_{\phi_o}$. To calculate the efficacy we need

$$\frac{\partial}{\partial\Delta} E_{\theta\phi}(T^p_N | G^*)\Big|_{\Delta=0} \quad \text{and} \quad \frac{\partial}{\partial\Delta} E_{\theta\phi}(T^p_N | G)\Big|_{\Delta=0}.$$

By the mean value theorem

$$H^*_{\theta\phi}(x) = G^*_\theta(x) + \frac{n_2}{N}(G^*_\phi(x) - G^*_\theta(x))$$

$$= G^*_\theta(x) - \frac{n_2\Delta}{N} \frac{\partial}{\partial\mu} G^*_\mu(x)\Big|_{\mu=\hat{\phi}} \quad \text{for some } \hat{\phi}.$$

$\theta \geq \hat{\phi} \geq \phi$. Thus

$$E_{\theta\phi}(T^p_N | G^*) = \int_{-\infty}^{\infty} J_p\left(G^*_\theta(x) - \frac{n_2\Delta}{N} \frac{\partial}{\partial\mu} G^*_\mu\Big|_{\mu=\hat{\phi}}\right) dG^*_\theta(x).$$

Another application of the mean value theorem yields

$$E_{\theta\phi}(T_N^p|G^*) = \int_{-\infty}^{\infty} J_p(G_\theta^*(x))dG_\theta^*(x)$$

$$- \frac{n_2\Delta}{N} \int_{-\infty}^{\infty} \frac{\partial}{\partial\mu} G_\mu^*(x)\Big|_{\mu=\hat{\phi}} J_p'(G_\mu^*(x))\Big|_{G_\mu^*=\hat{G}_\theta^*} dG_\theta^*(x)$$

where $\hat{G}_\theta^*$ lies between G_θ^* and $G_\theta^* - \frac{n_2}{N}\Delta \frac{\partial G_\mu^*}{\partial\mu}\Big|_{\mu=\hat{\phi}}$.

Recalling that the first integral has been shown to be zero, we have

$$\frac{\partial}{\partial\Delta} E_{\theta\phi}(T_N^p|G^*)\Big|_{\Delta=0} = -\left(\frac{n_2}{N}\right) \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} G_\theta^*(x)\right) J_p'(G_\theta^*(x))dG_\theta^*(x)\Big|_{\theta=\phi_o}.$$

Integrating by parts we have the latter equal to

$$-\left(\frac{n_2}{N}\right) \left| \frac{\partial}{\partial\theta} G_\theta^*(x)\Big|_{\phi_o} J_p \quad G_{\phi_o}^*(x) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} \frac{g_\theta^*(x)}{g_{\phi_o}(x)}\Big|_{\phi_o} J_p \quad G_{\phi_o}^*(x) \; dG_{\phi_o}^*(x) \right|$$

or

$$\frac{\partial}{\partial\Delta} E_{\theta\phi}(T_N^p|G^*)\Big|_{\Delta=0} = \left(\frac{n_2}{N}\right) \int_{-\infty}^{\infty} J_p^*(G_{\phi_o}^*)J_p(G_{\phi_o}^*)dG_{\phi_o}^*(x)$$

$$-\left(\frac{n_2}{N}\right) \left[\frac{\partial}{\partial\theta} G_\theta^*(x)\Big|_{\phi_o} J_p \quad (G_{\phi_o}^*(x))\Big|_{-\infty}^{\infty} \right]$$

We conclude

$$E_{T^p}(G^*) =$$

$$\frac{p\left\{\int_{-\infty}^{\infty} J_{G^*}(G^*_{\phi_o}) J_G(G^*_{\phi_o}) dG^*_{\phi_o}(x) - \left[\frac{\partial}{\partial\theta} G^*_\theta(x) J_G\, G^*_{\phi_o}(x)\Big|_{-\infty}^{\infty}\right]\right\} + (1-p)\inf G^*_{\phi_o}}{p^2 \inf G_{\phi_o} + 2p(1-p)\int J_G(x) J_{G^*}(x)dx + (1-p)\inf G^*_{\phi_o}}$$

Similarly

$$E_{T^p}(G^*) =$$

$$\frac{p \inf G_{\phi_o} + (1-p)\left\{\int_{-\infty}^{\infty} J_{G^*}(G_{\phi_o}) J_G(G_{\phi_o}) dG_{\phi_o}(x) - \left[\frac{\partial}{\partial\theta} G_\theta(x) J_{G^*}(G_{\phi_o}(x))\Big|_{-\infty}^{\infty}\right]\right\}}{p^2 \inf G_{\phi_o} + 2p(1-p)\int J_G(x) J_{G^*}(x)dx + (1-p)\inf G^*_{\phi_o}}$$

Thus the conditions (a), (b) and (c) are satisfied and we can calculate ARE's by taking ratios of efficacies. These calculations allow us to prove an interesting lemma which shows <u>surprising symmetry</u> relations among these tests.

LEMMA: If

$$\frac{\partial}{\partial\theta} G_\theta(x)\Big|_{\phi_o} \frac{\partial}{\partial\theta} \log g^*_\theta(x)\Big|_{\phi_o}\Big|_{-\infty}^{\infty} = \frac{\partial}{\partial\theta} G^*_\theta(x)\Big|_{\phi_o} \frac{\partial}{\partial\theta} \log g_\theta(x)\Big|_{\phi_o}\Big|_{-\infty}^{\infty}$$

then

$$E_{T'T^o}(G_{\phi_o}) = E_{T^oT'}(G^*_{\phi_o}).$$

Further if $\inf G_{\phi_o} = \inf G^*_{\phi_o}$ then

$$E_{T^p}(G^*) = E_{T^{1-p}}(G).$$

Proof: This result is immediate when one notices that

$$J_{G^*}(G_{\phi_o}(x)) \Big|_{-\infty}^{\infty} = J_{G^*}(1) - J_{G^*}(0)$$

$$= J_{G^*}(G_{\phi_o}(x)) \Big|_{-\infty}^{\infty} ,$$

$$J_G(G^*_{\phi_o}(x)) \Big|_{-\infty}^{\infty} = J_G(G_{\phi_o}(x)) \Big|_{-\infty}^{\infty}$$

and

$$\int_{-\infty}^{\infty} J_G(G_{\phi_o})J_{G^*}(G_{\phi_o})dG_{\phi_o} = \int_{-\infty}^{\infty} J_{G^*}(G^*_{\phi_o})J_G(G^*_{\phi_o})dG^*_{\phi_o} . \quad \text{Q. E. D.}$$

Example 1. Suppose Λ consists of two elements, $\lambda = 1$ denoting normal c.d.f.'s with unit variance and $\lambda = 2$ denoting logistic distributions

$$g_\theta(x) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2} \qquad -\infty < x < \infty.$$

Then the LMPRT's for the unknown location parameter are the

Fisher-Yates, C_N, and the Wilcoxin, W_N, respectively. We easily calculate:

$$\frac{\partial}{\partial\Delta} E(W_N | \text{Logistic}) = \frac{1}{3} \frac{n_2}{N}$$

$$\frac{\partial}{\partial\Delta} E(W_N | \text{Normal}) = \frac{1}{\sqrt{\pi}} \frac{n_2}{N}$$

$$\sigma^2_{W_N} = \frac{1}{3} \frac{n_2}{n_1 N}$$

$$\frac{\partial}{\partial\Delta} E(C_N | \text{Logistic}) = \frac{1}{\sqrt{\pi}} \frac{n_2}{N}$$

$$\frac{\partial}{\partial\Delta} E(C_N | \text{Normal}) = \frac{n_2}{N}$$

$$\sigma^2_{C_N} = \frac{n_2}{n_1 N}$$

Letting $\Phi(x)$ represent the standard normal c.d.f. we have

$$\begin{aligned} \int_0^1 J_W(x) J_{C_N}(x) dx &= \int_0^1 (2x-1)\Phi^{-1}(x)dx \\ &= \int_{-\infty}^{\infty} (2\Phi(y)-1)y d\Phi(y) \\ &= \int_{-\infty}^{\infty} (2\Phi(y)-1)y\Phi'(y)dy. \end{aligned}$$

Integrating by parts and noticing that

$$-y\Phi'(y) = \Phi''(y)$$

we have

$$\int_0^1 J_W(x) J_{C_N}(x)dx = 2\int_{-\infty}^{\infty} (\Phi'(y))^2 dy$$

$$= \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2} dy$$

$$= \frac{1}{\sqrt{\pi}}$$

Letting $T_N^p = pW_N + (1-p)C_N$ we find

$$E_{T_N^p, C_N}(\text{Normal}) = \frac{\left[\frac{p}{\sqrt{\pi}} + (1-p)\right]^2}{\frac{p^2}{3} + (1-p)^2 + \frac{2p(1-p)}{\sqrt{\pi}}}$$

$$E_{T_N^p, W_N}(\text{Logistic}) = \frac{\left[\frac{p}{3} + \frac{(1-p)}{\sqrt{\pi}}\right]^2}{\frac{p^2}{3} + (1-p)^2 + \frac{2p(1-p)}{\sqrt{\pi}}}$$

We may modify these distributions by a change of scale of one of them to give all cdf's in Ω the same information: Let the variance σ^2 of the normal random variable be equal to $\frac{1}{3}$. Then

$$E_{T_N^p, C_N}(\text{Normal}) = \frac{\left[p\sqrt{\frac{3}{\pi}} + (1-p)\right]^2}{\left[p^2 + (1-p)^2 + 2p(1-p)\sqrt{\frac{3}{\pi}}\right]}$$

$$E_{T_N^p, W_N}(\text{Logistic}) = \frac{\left[p + (1-p)\sqrt{\frac{3}{\pi}}\,\right]^2}{[p^2 + (1-p)^2 + 2p(1-p)\sqrt{\frac{3}{\pi}}\,]}$$

$$= E_{T^{1-p}, C_N}(\text{Normal}) \quad .$$

This example, together with Figure 4, illustrates the symmetry uncovered in the lemma.

Example 2. Suppose Ω again consists of two elements, $\lambda = 1$: normal cdf's with unit variance, and $\lambda = 2$: double exponential cdf's:

$$L(x,\Theta) = \begin{cases} \frac{1}{2}e^{(x-\Theta)} & x \leqq \Theta \\ 1 - \frac{1}{2}e^{(\Theta-x)} & x \geqq \Theta \end{cases}$$

with density function

$$\ell(x,\Theta) = \frac{1}{2}e^{-|x-\Theta|} \qquad -\infty < x < \infty \ .$$

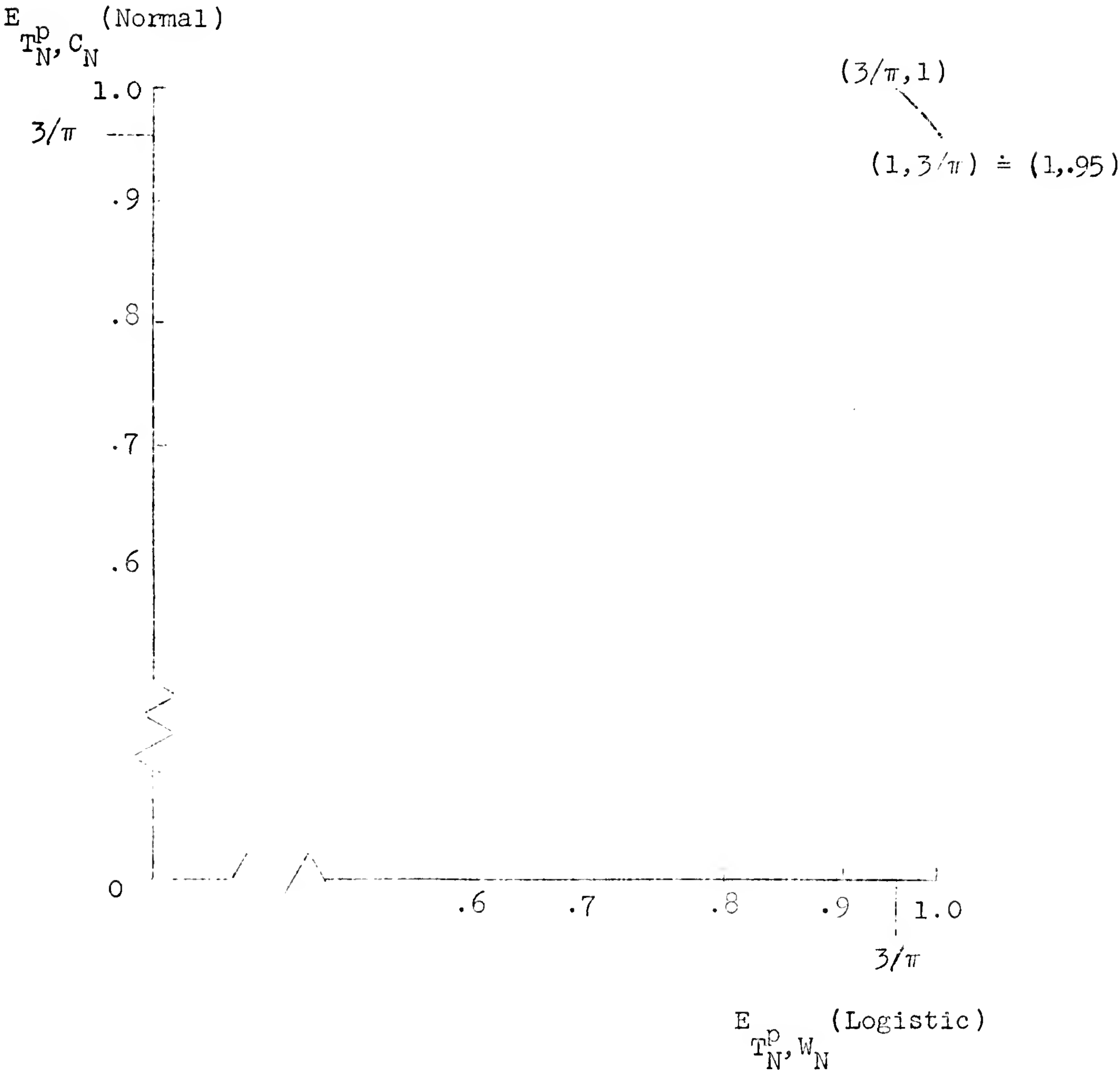
Notice that both distributions have unit information with regard to the unknown location parameter. The LMPRT's in this case are the Fisher-Yates, C_N, and the test given by Birnbaum (1962) and investigated by Laska (1962,a). We easily calculate:

$$\frac{\partial}{\partial\Delta} E(B_N | \text{double exponential}) = \frac{n_2}{N}$$

$$\frac{\partial}{\partial\Delta} E(B_N | \text{Normal}) = \sqrt{\frac{2}{\pi}}\,\frac{n_2}{N}$$

Figure 4

ASYMPTOTIC RISK POINTS OF $T_N^p = pW_N + (1-p)C_N$



$$\sigma^2_{B_N} = \frac{n_2}{n_1 N}$$

$$\frac{\partial}{\partial\Delta} E(C_N | \text{double exponential}) = \sqrt{\frac{2}{\pi}} \frac{n_2}{N}$$

$$\frac{\partial}{\partial\Delta} E(C_N | \text{Normal}) = \frac{n_2}{N}$$

$$\sigma^2_{C_N} = \frac{n_2}{n_1 N}$$

Also

$$\int_0^1 J_L(x) J_\phi(x) dx = -\int_{-\infty}^0 \Phi^{-1}(\Phi(x)) d\Phi(x) + \int_0^\infty \Phi^{-1}(\Phi(x)) d\Phi(x)$$

$$= 2\int_0^\infty x d\Phi(x)$$

$$= \sqrt{\frac{2}{\pi}}$$

Letting $T_N^p = pB_N + (1-p)C_N$ we have

$$E_{T_N^p, C_N}(\Phi) = \frac{\left[p\sqrt{\frac{2}{\pi}} + (1-p)\right]^2}{p^2 + (1-p)^2 + 2\sqrt{\frac{2}{\pi}}\, p(1-p)}$$

$$E_{T_N^p, B_N}(L) = \frac{\left[p + (1-p)\sqrt{\frac{2}{\pi}}\right]^2}{p^2 + (1-p)^2 + 2\sqrt{\frac{2}{\pi}}\, p(1-p)}$$

$$= E_{T_N^{1-p}, C_N}(\Phi) \,.$$

These calculations, together with Figure 5, provide a second illustration of the symmetry uncovered in the lemma.

Example 3. Let Λ contain the three shapes represented in Examples 1,2. Then the asymptotic risk points of the ARTs constitute a concave surface in the unit cube; Figures 4 and 5 represent edges of this surface lying in two faces of this cube. Some relevant statements about this surface follow immediately, and others can be found by determining a few points in the third edge analogous to these, and especially by computing several risk points in the interior of this surface such as that of

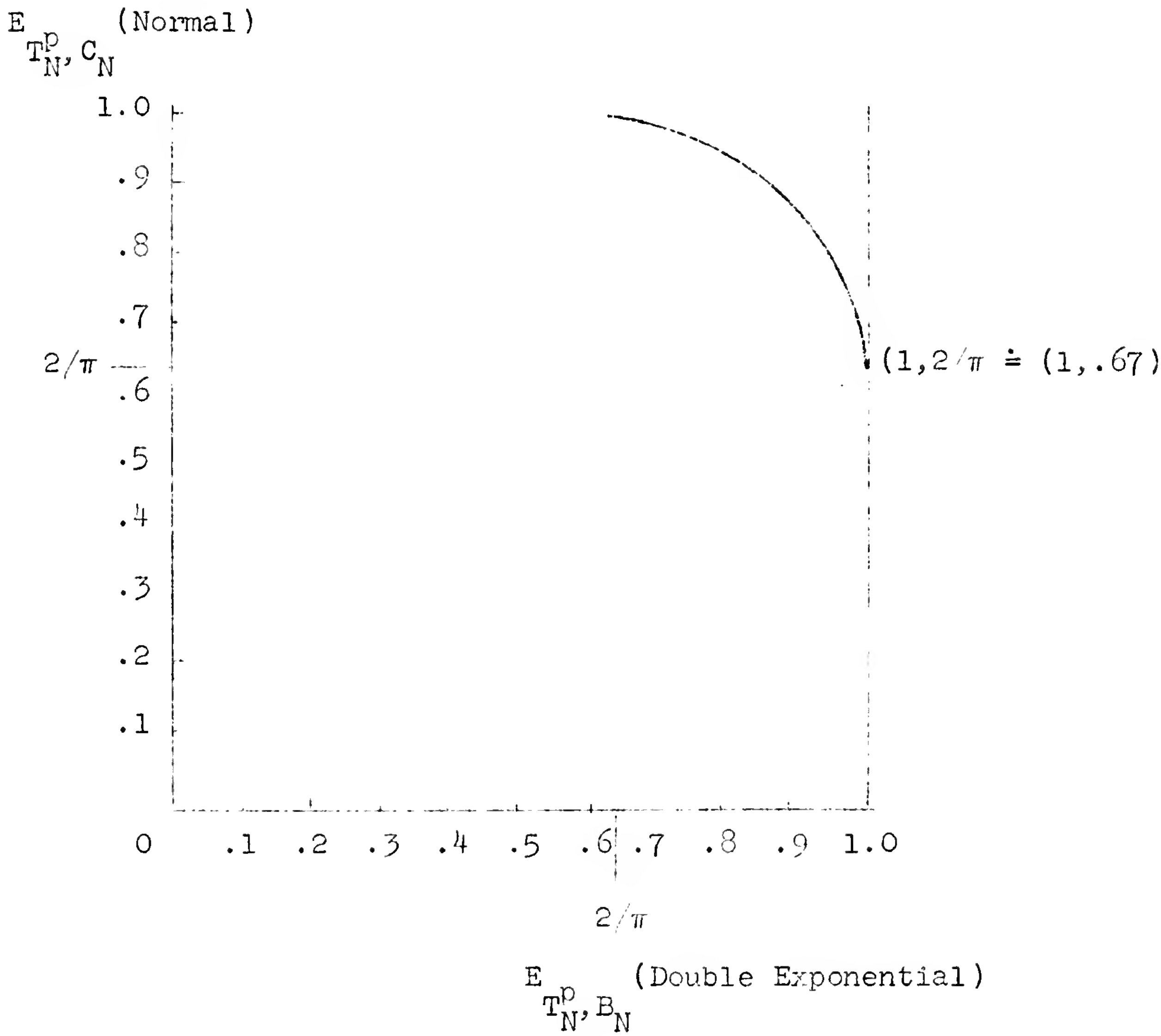
$$T_N = (1/3)C_N + (1/3)W_N + (1/3)B_N \ .$$

An interesting open problem is whether an ART is always also a LMPRT, under the assumption that Λ is a convex family of shapes in the sense defined in an analogous comment in Section 2 above.

3.4 Efficiency-robust estimators. The general method of constructing confidence limit estimators and median-unbiased point estimators from given testing methods is well known, and has been discussed recently by Hodges and Lehmann (1963) for the class of two-sample rank tests on which they base an approach to efficiency-robust estimation. The comments above on the practical and theoretical value of extending theory and techniques for LMPRTs to theory and techniques for ARTs, have direct analogues in connection with these problems of estimation. The actual construction of estimators from ARTs, and the interpretation of their efficiency

Figure 5

ASYMPTOTIC RISK POINTS OF $T_N^p = pB_N + (1-p)C_N$



properties, do not require special discussion since these consist of reinterpretation of test-efficiency properties in the context of corresponding estimation problems as in the reference above. It should not be overlooked that despite validity-robustness and efficiency-robustness properties which may be established for estimators based on LMPRTs or ARTs, there remains the important question whether such estimators or others have in addition robustness with respect to the strong assumption of identical form (apart from shift) of the two distributions. Conceivably the latter robustness problem could be approached by an extension of the above methods. Also to be considered in connection with the latter robustness question is the possibility of estimating shift by the difference of two efficiency-robust estimators of one of the kinds considered in Section 2 above.

4. Efficiency-robust unbiased estimators.

4.1 Introduction. Consider any family of density functions $f(x,\theta,\lambda)$, $\theta \in \Omega$, a subset of the real line, $\lambda \in \Lambda$. The problem of unbiased estimation of θ is that of finding estimators $t(x)$, satisfying $E(t(X)|\theta,\lambda) = \theta$ for each θ,λ, for which $\mathrm{var}(t|\theta,\lambda)$ takes values which are jointly suitably small in some specified sense for $\theta \in \Omega$, $\lambda \in \Lambda$.

The case in which λ is known (or in which no nuisance parameter λ appears), and in which the criterion of uniformly minimum variance (UMV) is adopted, is the only one which has been investigated very generally and systematically. In this case four

problems have been of principal interest:

1. Under what conditions on $f(x,\theta)$, θ admitting unbiased estimation for $\theta \in \Omega$, do UMV estimators exist? It has been shown (Rao (1945)), Blackwell (1947), that in case there exists a complete sufficient statistic, t, then the conditional expectation of s given t, where s is any unbiased estimator such that the variance of s is finite for each θ, will be UMV. In general, when no such statistic exists, there may or <u>may not be</u> a UMV estimator and no general theory now covers this case.

2. Under what further conditions are such UMV estimators essentially unique? It has been shown that existence of a complete sufficient statistic guarantees essential uniqueness. If none exists no general theory is available.

3. Under what conditions is a <u>given</u> estimator UMV? A sufficient condition is that it be unbiased and have variance equal identically to the Cramer-Rao lower bound; a strictly weaker sufficient condition is that the estimator be a function of complete sufficient statistic. But the latter condition is not necessary, and when it fails the problem is sometimes difficult and is not covered by available theory.

4. Can a randomized unbiased estimator be UMV? A negative answer follows immediately from the Rao-Blackwell Theorem.

For the wide class of problems in which no UMV estimator exists (for an estimable θ), the general goal of small values of $\text{var}(t|\theta)$ may be formulated as high precision at or near some specified value θ_0 of special interest; more precisely, a locally-best (LB) estimator has been defined as an unbiased estimator which

minimizes $var(t|\theta_o)$, where θ_o is a specified value of θ. Stein (1950) investigated the four problems above as they apply to LB rather than UMV estimators, giving conditions for existence and uniqueness, and a functional equation characterizing LB estimators that randomized estimators cannot be LB is seen in the same way as with UMV property above.

The third and most general sense of smallness of values of $var(t|\theta)$ has not previously been systematically studied. It is that of admissibility of an unbiased estimator t defined with reference to the variance function $var(t|\theta)$, $\theta \in \Omega$. Therefore we present here generalizations, using this criterion, of Stein's results. Among unbiased estimators, conditions are given for existence and uniqueness of Bayes solutions (with risk defined as variance); the uniqueness conditions automatically entail admissibility. A theorem on completeness of the class of Bayes solutions is given. These generalizations are presented along with their further generalization to the case of principal interest in the present paper, that in which an additional nuisance (or "shape") parameter λ is present: We deal then with densities $f(x,\theta,\lambda)$; unbiased estimators $t(x)$ satisfying $E(t|\theta,\lambda) = \theta$ for all $\theta \in \sum$ and $\lambda \in \wedge$; and admissibility of such estimators defined in the usual way with reference to their variance functions $var(t|\theta,\lambda)$.

No examples of applications of these generalizations are given; even for the special case of locally-best estimators, the present writers are not aware of any non-trivial applications to examples.

4.2 Uniqueness of Bayes solutions. In this section we are concerned with admissible unbiased estimators of a real valued parameter θ of the density function $f(x,\theta\lambda)$, $\theta \in \Sigma$, $\lambda \in \Lambda$, $\Sigma \times \Lambda = \Omega$ where x is a point in the σ-finite measure space $(R, \mathcal{A}, \mu)$. Let U be the convex set (i. e. $\alpha\delta_1 + (1-\alpha)\delta_2 \in U$ if $\delta_1, \delta_2 \in U$, for $0 \leqq \alpha \leqq 1$) of unbiased estimators of θ. We assume at the outset that U is not empty. We shall be interested in those $\delta^*(x) \in U$ for which

$$r^*(\delta^*, J) = \int_\Omega \int_R (\delta^*(x)-\theta)^2 f(x,\theta\lambda)d\mu(x)dJ(\theta,\lambda)$$

$$= \inf_{\delta \in U} \int_\Omega \int_R (\delta(x)-\theta)^2 f(x,\theta\lambda)d\mu(x)dJ(\theta,\lambda)$$

$$= \inf_{\delta \in U} \int_\Omega \sigma^2_{\theta,\lambda}(\delta)dJ(\theta,\lambda)$$

where $\sigma^2_{\theta,\lambda}(\delta) = E_{\theta,\lambda}[\delta(X)-\theta]^2$ and $J(\theta,\lambda)$ is any a priori measure over Ω. $\delta^*(x)$ is a Bayes solution with respect to J. The reason for this interest follows from the fact that if $\delta^*(x)$ is the unique estimator minimizing $\int_\Omega \sigma^2_{\theta,\lambda}(\delta)dJ(\theta\lambda)$ for any distribution function $J(\theta\lambda)$ over Ω, then $\delta^*(x)$ is admissible.

Define $\int_\Omega f(x,\theta\lambda)dJ(\theta\lambda) = f_J(x)$. We now give a condition for uniqueness of Bayes solutions:

LEMMA 4.1. If $\int_R \frac{[f(x,\theta\lambda)]^2}{f_J(x)} d\mu(x) < \infty$ for all $(\theta,\lambda) \in \Omega$ and if for some $\delta \in U$ the Bayes risk $r^*(\delta, J)$ is finite, then there

exists $\delta^*(x)$ such that

$$\int_\Omega \sigma^2_{\theta,\lambda}(\delta^*)dJ(\theta\lambda) = \inf_{\delta \in U} \int_\Omega \sigma^2_{\theta,\lambda}(\delta)dJ(\theta\lambda)$$

Moreover δ^* is essentially unique, i. e., up to an almost everywhere (μ) equivalence.

Proof: For any $\delta \in U$ the risk is given by

$$r^*(\delta,J) = \int_\Omega\int_R (\delta(x)-\theta)^2 f(x,\theta\lambda)d\mu(x)dJ(\theta\lambda)$$

$$= \int_\Omega\int_R \delta^2(x)f(x,\theta\lambda)d\mu(x)dJ(\theta\lambda) - \int_\Omega \theta^2 dJ(\theta\lambda) < \infty .$$

The problem is thus equivalent to minimizing

$$\int_\Omega\int_R \delta^2(x)f(x,\theta\lambda)d\mu(x)dJ(\theta\lambda) = \int_R \delta^2(x)f_J(x)d\mu(x)$$

subject to

$$\int_R \delta(x)f(x,\theta\lambda)d\mu(x) \equiv \theta$$

for all $(\theta,\lambda) \in \Omega$.

The remainder of the proof is identical with that of Stein (1950) who considered the case corresponding to an a priori distribution over Ω with all of its mass at one point $\theta_o\lambda_o$. Stein's other results may also be extended to gain an interesting and useful

description of admissible unbiased estimators. We shall illustrate by stating a generalization of his "Principal Theorem".

<u>THEOREM 4.1</u> (Generalization of Stein's "Principal Theorem"): Suppose that for an <u>a priori</u> distribution $J(\theta\lambda)$ over Ω, $\int_R \frac{[f(x,\theta\lambda)]^2}{f_J(x)} d\mu(x) < \infty$, $\frac{f(x,\theta\lambda)}{f_J(x)}$ is finite for all $(\theta,\lambda) \in \Omega$ and almost all x and that there exists an unbiased estimator δ of θ for which $\int \sigma^2_{\theta,\lambda}(\delta) dJ(\theta\lambda) < \infty$. Then δ^* is the Bayes solution (essentially unique by virtue of Lemma 4.1) with respect to J if and only if there is a real-valued functional T over the set G of functions of the form

$$\psi(\theta,\lambda) = \int_R \phi(x) f(x,\theta\lambda) d\mu(x)$$

where $\phi(x)$ satisfies

$$\int \phi^2(x) f_J(x) d\mu(x) < \infty$$

such that

$$T\left(\int \frac{f(x,\theta_1,\lambda)}{f_J(x)} f(x,\theta\lambda) d\mu(x) \right) = \theta_1$$

for all $(\theta,\lambda) \in \Omega$ and

$$T\left(\int \phi(x) f(x,\theta\lambda) d\mu(x) \right) = \int \phi(x) \delta^*(x) f_J(x) d\mu(x).$$

Stein's corollaries (pp. 409, 410) can also be generalized in a natural manner to provide explicit methods of determination of T and therefore of the Bayes solutions which, at the very least, form an important subset of the admissible class.

4.3 <u>A complete class theorem</u>. The main content of this section is Theorem 4.2 which asserts that under certain conditions the class of Bayes solutions is essentially complete. We begin with a few technical lemmas.

LEMMA 4.2. Let $(\mathscr{R}, \mathscr{A}, \mu)$ be a σ-finite measure space and $(\delta_n;\ n = 0,1,2\ldots)$ be a sequence of $\mathscr{A}$-measurable functions such that

$$\int_B \delta_n d\mu \to \int_B \delta_o d\mu$$

for all $B \in \mathscr{A}$ with $\mu(B) < \infty$. Suppose $|\delta_n(x)| \leqq C$ for all $n = 0,1,2,\ldots$ $x \in R$, where C is a fixed but arbitrary constant. Then

$$\int_R \delta_n g d\mu \to \int_R \delta_o g d\mu$$

for all $g \in L^1(\mu)$, the space of absolutely integrable functions with respect to the measure μ.

For proof see Dunford and Schwartz (1961, Problem 6, p. 339).

LEMMA 4.3 Let Ω be a compact metric space and $(\pi_n;\ n = 0,1,2,\ldots)$ be a sequence of probability measures on Ω. In order that

$$\int_\Omega g(\theta)d\pi_n(\theta) \to \int_\Omega g(\theta)d\pi_o(\theta)$$

for all continuous g on Ω it is sufficient that for some covering net $[D_{k_1 \ldots k_m}]$

$$\pi_n(\bar{D}_{k_1 \ldots k_m}) \to \pi_o(\bar{D}_{k_1 \ldots k_m}) \qquad \text{for all } k_1 \ldots k_m .$$

(For the definition of a covering net see Wald (1950, p. 66.))

LEMMA 4.4. Let $\delta(A, x)$ be a probability measure in A for each x and measurable for each $A (A \subseteq \Omega,\ x \subseteq R)$. Let $g(\theta')$ be a bounded measurable function of $\theta' \subseteq \Omega$. Define (for γ any probability measure on R)

$$\nu(A) = \int_R \delta(A, x) d\gamma(x) .$$

Then ν is a probability measure on Ω and

$$\int_\Omega g(\theta') d\nu(\theta') = \int_R \left(\int_\Omega g(\theta') d\delta(\theta', x) \right) d\gamma(x).$$

<u>Note</u>: Ω being a compact metric space implies that the Borel sets form a σ-algebra.
For proof see Loeve (1955, Problem 13, p. 140).

<u>THEOREM 4.2</u>. Let $\sum$ be a compact subset of the real line, and $f(x, \theta)$ continuous in θ. Then the class of Bayes solutions with respect to <u>a priori</u> distributions $G(\theta)$ over $\sum$ is essentially complete where the space of decision functions U_r is the class of unbiased, <u>randomized</u> estimators, $\delta(x, A)$. (That is, $\delta(x, A)$ is

a probability measure over $\underline{\sum}$ for each x and measurable for each measurable set A, $A \subseteq \underline{\sum}$, $x \in R$.) Further, if $\underline{\sum}$ is convex then any purely randomized decision function is inadmissible.

Proof: It is sufficient to show that the assumptions of Wald (1950, 3.1 to 3.7) are satisfied. Except for the requirement that the class of decision functions be closed in the sense of "regular convergence"(Wald (1950, Assumption 3.6 (ii), p. 68)), the verification of the other requirements is analogous to that in the well-known case of mean squared error loss function without restriction to unbiasedness. Let $\delta(x)$ be a member of the class U_r of randomized, unbiased estimators, i. e., for each x, $\delta(x)$ is a probability measure over Ω. Let $\overline{D}$ be any measurable subset of Ω. For fixed x and a given decision function δ, let us denote the probability that the decision lies in $\overline{D}$ by $\delta(x,\overline{D}) = \int_{\overline{D}} d\delta(x,\theta)$. To say $\delta(x)$ is unbiased is to require that

$$\int_R \left[\int_{\underline{\underline{\quad}}} \theta' d\delta(x,\theta') \right] f(x,\theta) d\mu(x) \equiv \theta \qquad \text{for all} \quad \theta \in \underline{\sum} .$$

We now show that the class of randomized estimators, U_r is closed in the sense of regular convergence. Let $[\overline{D}_{k_1 k_2 \ldots k_m}]$ be a covering net of $\underline{\sum}$ and suppose that

$$\lim_{n \to \infty} \int_{R_S} \delta_n(\overline{D}_{k_1 \ldots k_m}, x) d\mu(x) = \int_{R_S} \delta(\overline{D}_{k_1 \ldots k_m}, x) d\mu(x)$$

for every $\overline{D}_{k_1 \ldots k_m}$ in the net and every set $R_S \in \mathcal{V}$ of finite μ

measure, where $[\delta_n(x)]$ is a sequence of randomized unbiased estimators. Then we must show that $\delta(x)$ is a member of U_r.

Let $\eta = \left\{\overline{D}_{k_1 \ldots k_m}\right\}$.

From Lemma 4.2, with $c = 1$, we get

$$\int_R \delta_n(A,x)g(x)d\mu(x) \to \int_R \delta_o(A,x)g(x)d\mu(x)$$

for all $g \in L_1(\mu)$ and $A \in \eta$.

In particular taking $g(x) = f(x,\theta)$ we have

$$\int_R \delta_n(A,x)f(x,\theta)d\mu(x) \to \int_R \delta_o(A,x)f(x,\theta)d\mu(x)$$

for all θ and all $A \in \eta$. Write

$$dP_\theta = f(x,\theta)d\mu$$

and

$$v_n^\theta(A) = \int_R \delta_n(A,x)dP_\theta(x) .$$

Rewriting the above we have for each arbitrary but fixed θ

$$v_n^\theta(A) \to v_o^\theta(A) \qquad \text{for all } A \in \eta .$$

Hence, by Lemma 4.3 for any continuous function $g(\theta')$ over $\sum$

$$\int_{\sum} g(\theta')dv_n^\theta(\theta') \to \int_{\sum} g(\theta')dv_o^\theta(\theta')$$

(where in Lemma 4.3 we have let $\pi_n = v_n^{\theta}$).

Now $v_n^{\theta}(A) = \int_R \delta_n(A,x)dP_{\theta}(x)$. Letting $\gamma = P_{\theta}$ in Lemma 4.4 yields

$$\int_{\sum} g(\theta')dv_n^{\theta}(\theta') = \int_R \left(\int_{\sum} g(\theta')d\delta_n(\theta',x)\right) dP_{\theta}(x) .$$

Therefore

$$\int_R \int_{\sum} g(\theta')d\delta_n(\theta',x) \quad dP_{\theta}(x) \to \int_R \int_{\sum} g(\theta')d\delta_o(\theta',x) \quad dP_{\theta}$$

Writing $g(\theta') = \theta'$ for all $\theta' \in \sum$ gives

$$\int_R \int_{\sum} \theta' d\delta(x,\theta')f(x,\theta)d\mu(x) = \theta \qquad \text{for all} \quad \theta \in \sum$$

and $\delta(x)$ is an unbiased randomized estimator. Thus U_r is closed in the sense of regular convergence and the assumptions of Wald are satisfied. Therefore essential completeness is proved.

Frequently it will not be necessary to use a randomized estimator for one may apply the Rao-Blackwell theorem, (Rao (1945), Blackwell (1947)). That is, for any fixed x, $\delta(x)$ defines a probability measure over $\sum$. Consider for use as an estimator the conditional expected value of θ, under this distribution, given x. If $\sum$ is convex this new non-randomized estimator is unbiased, it has variance less than or equal to that of its generator, the randomized estimator, which is therefore inadmissible.

The proof is now complete.

An extension of this theorem to the case pertinent to robustness is immediate. Introduction of the "nuisance" parameter λ disturbs only the property that Wald's assumption 3.7 (Wald 1950, p. 96) is satisfied. Let

$$V = \left\{ f(x,\theta,\lambda) \right\} \lambda \in \Lambda, \quad \theta \in \Sigma \quad \text{and let} \quad f_n \in V \; ; \quad n = 1,2,3\ldots$$

and

$$t_r(f_n, f) = \sup_{M \subset R} \left| \int_M [f_n(x,\theta,\lambda) - f(x,\theta,\lambda)] d\mu(x) \right|$$

where M is any μ-measurable subset of R. Assumption 3.7 concerns itself with compactness of V in the sense of the metric t_r and the property that if $f_n \to f$ in the sense of t_r, then

$$\sup_{\hat{\theta} \in \Sigma} \left| W(f_{n_0}, \hat{\theta}) - W(f, \hat{\theta}) \right| \to 0$$

where W is the loss function $(\theta-\hat{\theta})^2$. Since assumptions 3.1 to 3.6 are not affected by λ we have a

COROLLARY. If $f(x,\theta,\lambda)$ is a family of density functions, $\lambda \in \Lambda$ where Λ contains finitely many points such that for each fixed λ the assumptions of theorem 4.2 are satisfied, then the class of Bayes solutions with respect to <u>a priori</u> distributions over

$\sum \times \Lambda$ is essentially complete with respect to the class of randomized unbiased estimators.

The proof of the corollary follows immediately from the fact that the assumptions 3.1-3.9 are satisfied in the same way as in the case when Λ contained only one point. We summarize with the following corollary.

COROLLARY. If for each J the conditions of Lemma 4.1 and the conditions of Theorem 4.2 are satisfied and if Λ contains finitely many points then the class of Bayes solutions forms an essentially complete class of admissible estimators.

References

1. Ansari, A. R. and R. A. Bradley. Rank-sum test for dispersions. Annals of Math. Stat., 31 (1960) pp. 1174-1189.
2. Birnbaum, Allan. Some theory and techniques for robust estimation. (Abstract) Annals of Math. Stat., 32 (1961) p. 622.
3. Birnbaum, Allan. Two-sample rank tests: (A) efficient against double-exponential alternatives; (B) efficiency-robust. (Abstract) Annals of Math. Stat., 33 (1962) p. 1211.
4. Birnbaum, Allan and Jack Dudman. Logistic order statistics. Annals of Math. Stat., 34 (1963) p. 658-663.
5. Blackwell, D. Conditional expectation and unbiased sequential estimation. Annals of Math. Stat., 17 (1947) pp. 13-23.
6. Blom, Gunnar. Statistical estimates and transformed beta-variables. Wiley, New York, 1958.
7. Capon, Jack. Asymptotic efficiency of certain locally most powerful rank tests. Annals of Math. Stat., 32 (1961) pp. 88-100.
8. Chernoff, Herman and I. Richard Savage. Asymptotic normality and efficiency of certain nonparametric test statistics. Annals of Math. Stat., 29 (1958) p. 972-994.
9. Dunford, N. and J. T. Schwartz. Linear Operators. Interscience Publishers, New York, 1961.
10. Greenberg, B. and A. E. Sarhan (eds.). Contributions to order statistics. Wiley, New York, 1962.

11. Hastings, Cecil, Jr., Frederick Mosteller, John Tukey and Charles P. Winsor. Low moments for small samples: a comparative study of order statistics. Annals of Math. Stat., 18 (1947) pp. 413-126.

12. Hodges, J. L. and E. L. Lehmann. Estimates of location based on rank tests. Annals of Math. Stat., 34 (1963) pp. 598-611.

13. Huber, Peter J. Robust estimation of a location parameter. Annals of Math. Stat., 35 (1964) pp. 73-101.

14. Laska, Eugene M. A general theory of robustness. Ph. D. thesis, Department of Mathematics, New York University, 1962(a)

15. Laska, Eugene M. Complete class theorems for unbiased estimation. Courant Institute of Mathematical Sciences, New York University, Technical Report IMM-NYU 302, 1962(b).

16. Laska, Eugene M. A locally most powerful rank test for the location parameter of a double exponential distribution. Courant Institute of Mathematical Sciences, New York University, Technical Report IMM-NYU 304, 1962(c).

17. Lloyd, E. H. Least-squares estimation of location and scale parameters using order statistics. Biometrika, 39 (1952) pp. 88-95.

18. Loeve, M. Probability theory. D. Van Nostrand Co., Inc., New York, 1955.

19. Mikulski, P. W. On the efficiency of optimal nonparametric procedures in the two-sample case. Annals of Math Stat., 34 (1963) pp. 22-32.

20. Noether, Gottfried E. On a theorem of Pitman. Annals of Math. Stat., 26 (1955) pp. 64-68.

21. Pitman, E. J. G. Lecture notes on nonparametric statistical inference. Department of Mathematical Statistics, Columbia University, 1948.

22. Rao, C. R. Information and accuracy obtainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37 (1945) pp. 81-91.

23. Sarhan, Ahmed E. Estimation of the mean and standard deviation by order statistics. Annals of Math. Stat., 25 (1954) pp. 317-328.

24. Sarhan, Ahmed E. and Bernard G. Greenberg. Contributions to order statistics. Wiley, New York, 1962.

25. Stein, Charles. Unbiased estimators with minimum variance. Annals of Math. Stat., 21 (1950) pp. 406-415.

26. Tukey, John W. A survey of sampling from contaminated distributions. Paper 39 (pp. 448-485) in Contributions to probability and statistics (I. Olkin, et al, eds.) Stanford Univ. Press, 1960.

27. Tukey, John W. The future of data analysis. Annals of Math. Stat., 33 (1962) pp. 1-67.

28. Wald, A. Statistical decision functions. Wiley, New York, 1950.

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Courant Institute of Mathematical Sciences New York University	2a. REPORT SECURITY CLASSIFICATION Not classified
	2b. GROUP

3. REPORT TITLE "Optimal Robustness for Estimators and Tests"

4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical report

5. AUTHOR(S) (Last name, first name, initial)

Birnbaum, Allan and Laska, Eugene

6. REPORT DATE February 1965	7a. TOTAL NO. OF PAGES 55	7b. NO. OF REFS 28

8a. CONTRACT OR GRANT NO. NONR 285(38)	9a. ORIGINATOR'S REPORT NUMBER(S) IMM-NYU 334
b. PROJECT NO. 042-206	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None
d.	

10. AVAILABILITY/LIMITATION NOTICES "Qualified requesters may obtain copies of this report from DDC"

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY U S Navy, Office of Naval Research 207 West 24 St., New York, N.Y.

13. ABSTRACT Turkey (1960-1962) has provided a broad perspective for research in efficiency-robustness of estimators, as well as an important part of the knowledge available in this area. The present paper is intended to complement these by supplying formulations of concepts, techniques, and initial results for optimally efficiency-robust estimators and tests in several types of problems. Relations to Turkey's investigation are discussed in Section 2, with brief reference to the related work of Huber (1964). Relations to the approach to robust estimation of Hodges and Lehmann (1963) are discussed in Section 3.

The present approach may be described as a formal indexing of alternative specifications (e.g. "shapes" of error-distributions) by a nuisance parameter, and adaptation of admissibility and related concepts and Bayes techniques of the Neyman-Pearson and Wald theories to the estimation and testing problems thus formulated. Specific problems for which new optimal efficiency-robust estimators are given are: linear estimation of location parameters (Section 2); rank tests and related estimators for two-sample problems (Section 3); and unbiased estimation (Section 4). A by-product included in Section 4 is a generalization of Stein's (1950) characterization of locally-best unbiased estimators to the class of admissible unbiased estimators together with the corresponding complete class theorem.

DD FORM 1 JAN 64 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

Distribution List for NONR 285 (38)

Head, Statistics Branch
Office of Naval Research
Washington 25, D.C. (2)

Commanding Officer
Office of Naval Research
Branch Office
Navy #100
Fleet Post Office
New York, New York (2)

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314 (20)

Technical Information Officer
Naval Research Laboratory
Washington 25, D.C. (6)

Institute for Defense Analyses
Communications Research Division
von Neumann Hall
Princeton, New Jersey (1)

Bureau of Supplies and Accounts
Code OW
Department of the Navy
Washington 25, D.C. (1)

Dr. Herbert Solomon
Department of Statistics
Stanford University
Stanford, California (1)

Professor Frank Spitzer
Department of Mathematics
Cornell University
Ithaca, New York (1)

Professor S.S. Wilks
Department of Mathematics
Princeton University (1)
Princeton, New Jersey

Professor Evan J. Williams
Institute of Statistics
State College Section
North Carolina State College
Raleigh, North Carolina (1)

Professor J. Wolfowitz
Department of Mathematics
Cornell University
Ithaca, New York (1)

Professor W.H. Kruskal
Department of Statistics
The University of Chicago
Chicago 37, Illinois (1)

Prof. T.W. Anderson
Dept. of Math. Statistics
Columbia University
New York 27, New York (1)

Prof. Z.W. Birnbaum
Laboratory of Stat. Research
Department of Mathematics
University of Washington
Seattle 5, Washington (1)

Prof. A.H. Bowker, Chancellor
City University of New York
535 East 80th Street
New York, New York 10021 (1)

Prof. Ralph A. Bradley
Florida State University
Department of Statistics
Tallahassee, Florida (1)

Prof. Herman Chernoff
Applied Math. and Stat. Lab.
Stanford University
Stanford, California (1)

Prof. W.G. Cochran
Department of Statistics
Harvard University
Cambridge, Massachusetts (1)

Prof. Benjamin Epstein
768 Garland Drive
Palo Alto, California (1)

Dr. Paul G. Hoel
Department of Mathematics
University of California
Los Angeles 24, California (1)

Prof. Harold Hotelling
Associate Director
Institute of Statistics
University of North Carolina
Chapel Hill, North Carolina (1)

Prof. L. Hurwicz
School of Business Administration
University of Minnesota
Minneapolis, Minnesota (1)

Prof. Leo Katz
Department of Statistics
Michigan State University
East Lansing, Michigan (1)

Prof. Oscar Kempthorne
Statistics Laboratory
Iowa State College
Ames, Iowa (1)

Dr. Karl F. Kossack
Director
Laboratory for Computer Sciences
P. O. Box 30365
Dallas, Texas 75230 (1)

Prof. Gerald J. Lieberman
Applied Math. and Stat. Laboratory
Stanford University
Stanford, California (1)

Prof. William G. Madow
Stanford Research Institute
Menlo Park, California (1)

Prof. J. Neyman
Department of Statistics
University of California
Berkeley 4, California (1)

Prof. Herbert Robbins
Mathematical Statistics Dept.
Columbia University
New York 27, New York (1)

Prof. Murray Rosenblatt
Department of Mathematics
Brown University
Providence 12, Rhode Island (1)

Prof. L.J. Savage
Department of Mathematics
Yale University
New Haven, Connecticut (1)

Prof. Howard Raiffa
Grad. School of Business Admin.
Harvard University
Cambridge, Massachusetts (1)

Prof. F. Mosteller
Department of Statistics
Harvard University
Cambridge, Massachusetts (1)

Prof. John Tukey
Department of Mathematics
Princeton University
Princeton, New Jersey (1)

Dr. Jerome Cornfield
National Heart Institute
Bethesda, Maryland (1)

Dr. Theodore Harris
The Rand Corporation
1700 Main Street
Santa Monica, California (1)

Prof. Borge Jessen
Matematiske Institut
University of Copenhagen
Blegdamsvej 15
Copenhagen, Denmark (1)

Prof. A. Charnes
The Technological Institute
Northwestern University
Evanston, Illinois (1)

Prof. Milton Sobel
Department of Mathematics
University of Minnesota
Minneapolis, Minnesota (1)

Prof. Agnes Berger
Department of Biostatistics
School of Public Health
Columbia University
Morningside Heights
New York, New York (1)

Prof. J. Hemelryjk
Matematische Centrum
Amsterdam, Holland (1)

Prof. Joseph Moyal
Canberra University
Canberra City, Australia (1)

Prof. Emanuel Parzen
Department of Statistics
Stanford University
Stanford, California (1)

Prof. Donald Darling
Department of Mathematics
University of Michigan
Ann Arbor, Michigan (1)

Prof. David Blackwell
Department of Statistics
University of California
Berkeley, California (1)

Prof. J. Kiefer
Cornell University
Ithaca, New York (1)

Prof. Leo Goodman
Department of Statistics
University of Chicago
Chicago 37, Illinois (1)

Dr. George B. Dantzig
Department of Mathematics
University of California
Berkeley, California (1)

Prof. Victor Klee
Department of Mathematics
University of Washington
Seattle, Washington (1)

Prof. Samuel Karlin
Mathematics Department
Stanford University
Stanford, California (1)

Prof. Eugene Lukacs
Institut Statistique
de l'Universite de Paris
Paris 5, France (1)

Dr. Elza Berquo
Caixa Postal 8099
Faculdade de Higiene
Sao Paulo, Brazil (1)

Prof. A. Dvoretsky
Department of Mathematics
Hebrew University
Jerusalem, Israel (1)

I.M.S.

Prof. Allan Birnbaum
25 Waverly Place (1)

Prof. Madan L. Puri
25 Waverly Place (1)

Prof. Warren [illegible] Hirsch
25 Waverly Place (1)

Decision Studies Group
460 California Avenue
Palo Alto, California
Att: Warren R. Ketler (1)

DATE DUE

MAY 6 67			
11·10·62			
NOV 5 1977			

N.Y.U. Courant Institute of Mathematical Sciences
251 Mercer St.
New York 12, N. Y.